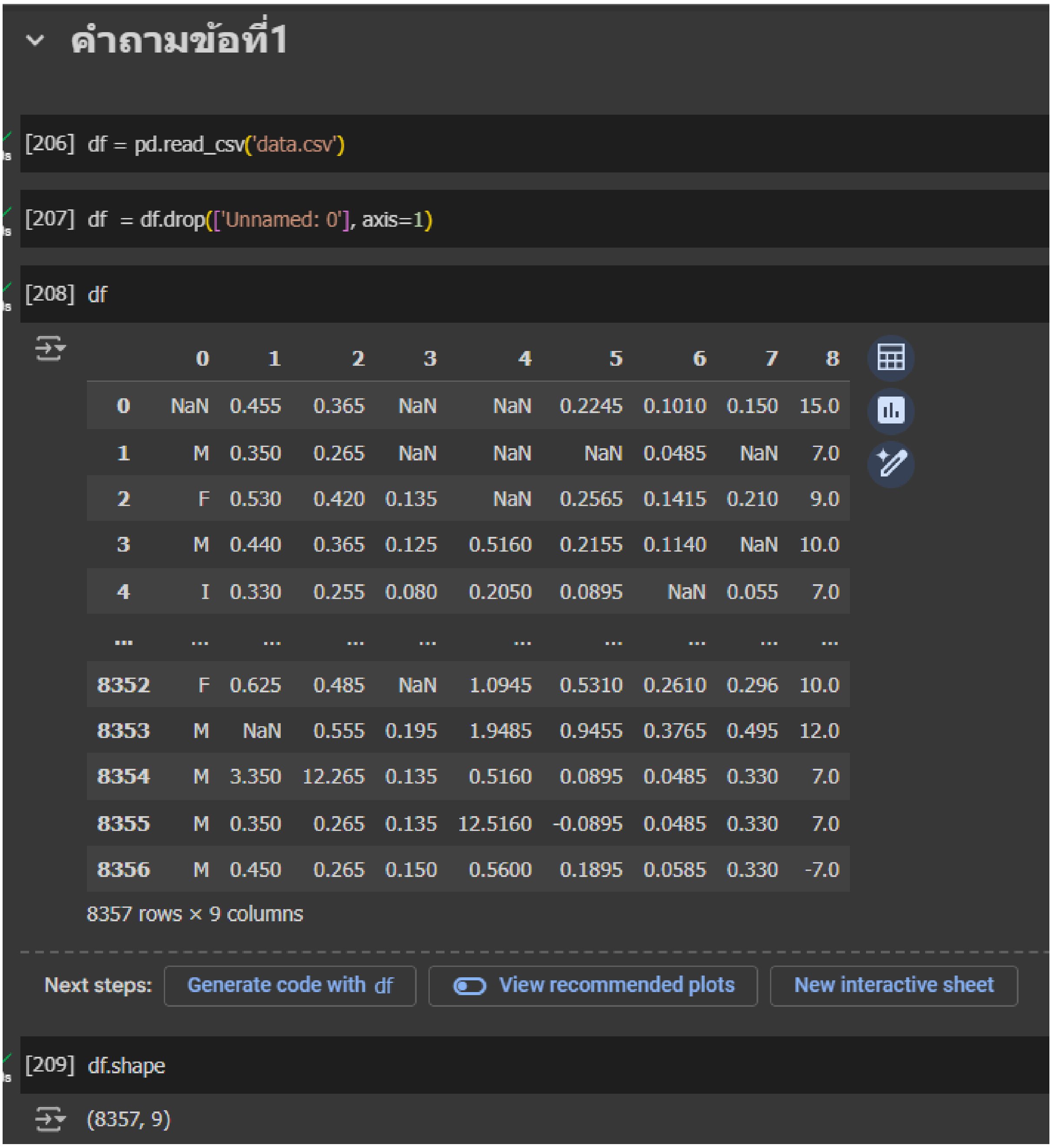
คำถามข้อที่ 1



คำถามข้อที่1

```
[206] df = pd.read_csv('data.csv')
```

```
[207] df  = df.drop(['Unnamed: 0'], axis=1)
```

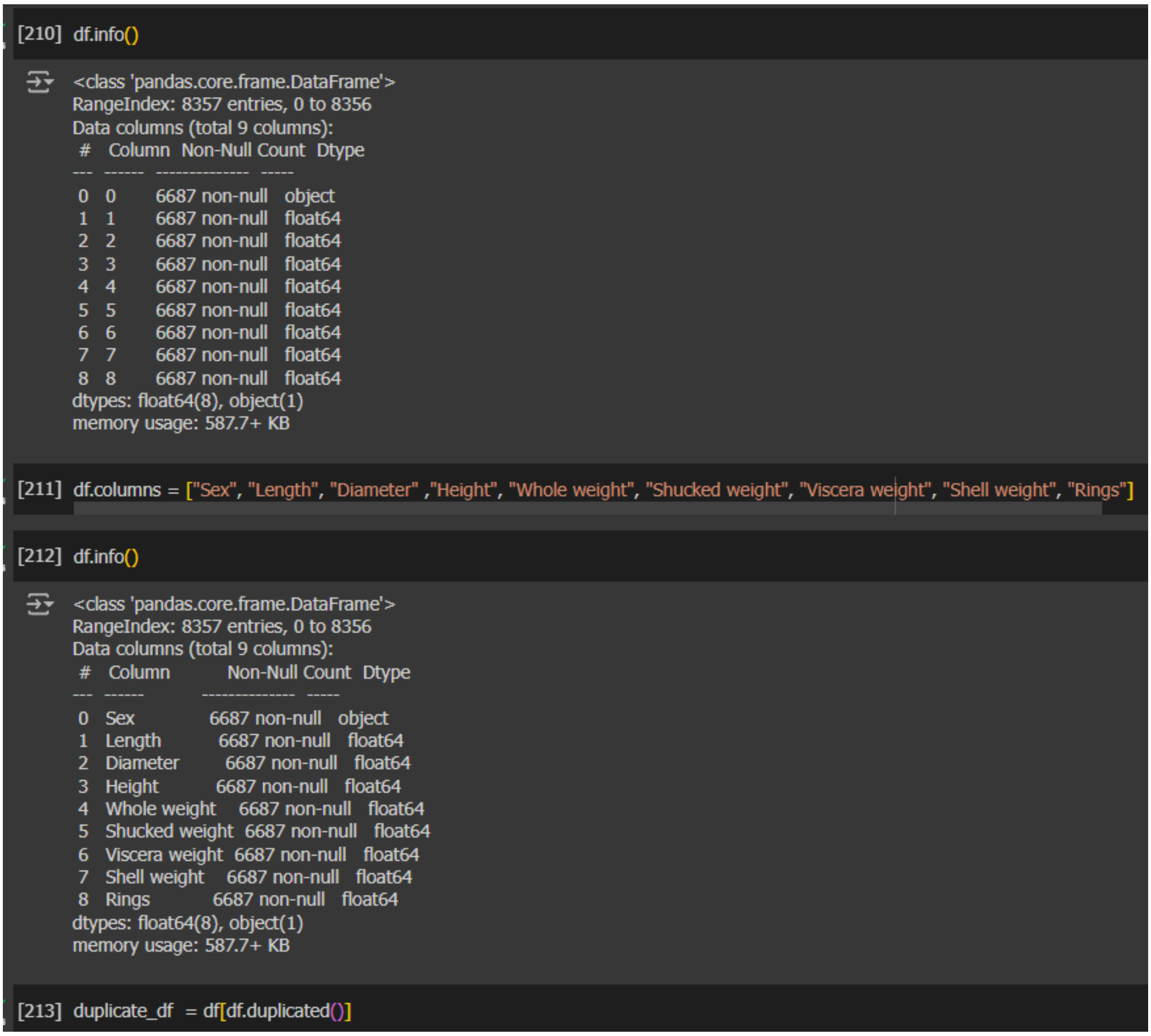
```
[208] df
```

	0	1	2	3	4	5	6	7	8
0	NaN	0.455	0.365	NaN	NaN	0.2245	0.1010	0.150	15.0
1	M	0.350	0.265	NaN	NaN	NaN	0.0485	NaN	7.0
2	F	0.530	0.420	0.135	NaN	0.2565	0.1415	0.210	9.0
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	NaN	10.0
4	I	0.330	0.255	0.080	0.2050	0.0895	NaN	0.055	7.0
...	...	...	...	...	...	...	...	...	...
8352	F	0.625	0.485	NaN	1.0945	0.5310	0.2610	0.296	10.0
8353	M	NaN	0.555	0.195	1.9485	0.9455	0.3765	0.495	12.0
8354	M	3.350	12.265	0.135	0.5160	0.0895	0.0485	0.330	7.0
8355	M	0.350	0.265	0.135	12.5160	-0.0895	0.0485	0.330	7.0
8356	M	0.450	0.265	0.150	0.5600	0.1895	0.0585	0.330	-7.0

8357 rows × 9 columns

```
[209] df.shape
```

```
(8357, 9)
```



```
[210] df.info()
```

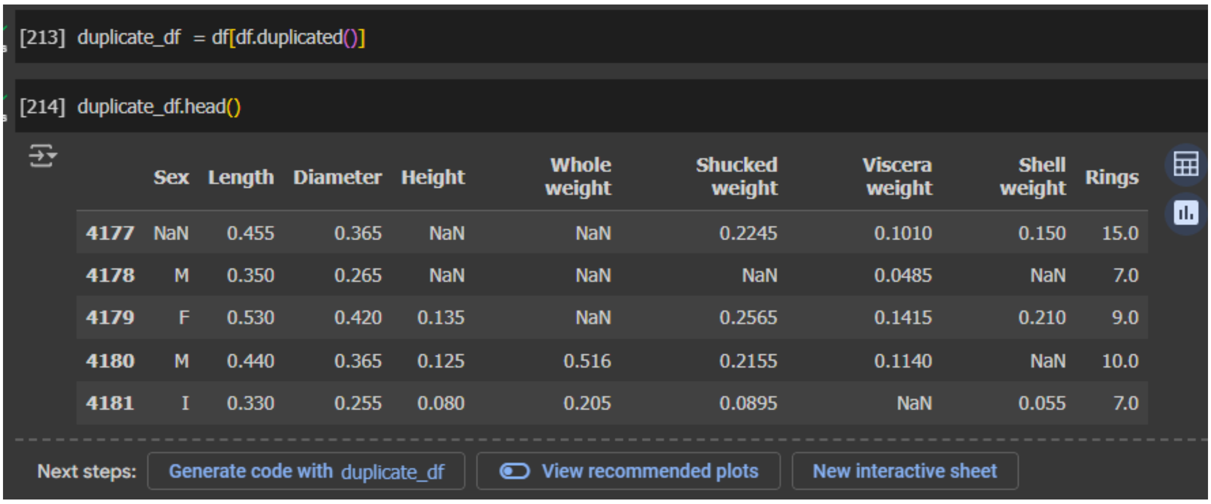
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8357 entries, 0 to 8356
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   0       6687 non-null   object
 1   1       6687 non-null   float64
 2   2       6687 non-null   float64
 3   3       6687 non-null   float64
 4   4       6687 non-null   float64
 5   5       6687 non-null   float64
 6   6       6687 non-null   float64
 7   7       6687 non-null   float64
 8   8       6687 non-null   float64
dtypes: float64(8), object(1)
memory usage: 587.7+ KB
```

```
[211] df.columns = ["Sex", "Length", "Diameter" ,"Height", "Whole weight", "Shucked weight", "Viscera weight", "Shell weight", "Rings"]
```

```
[212] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8357 entries, 0 to 8356
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Sex             6687 non-null   object
 1   Length          6687 non-null   float64
 2   Diameter        6687 non-null   float64
 3   Height          6687 non-null   float64
 4   Whole weight    6687 non-null   float64
 5   Shucked weight  6687 non-null   float64
 6   Viscera weight  6687 non-null   float64
 7   Shell weight    6687 non-null   float64
 8   Rings           6687 non-null   float64
dtypes: float64(8), object(1)
memory usage: 587.7+ KB
```

```
[213] duplicate_df  = df[df.duplicated()]
```

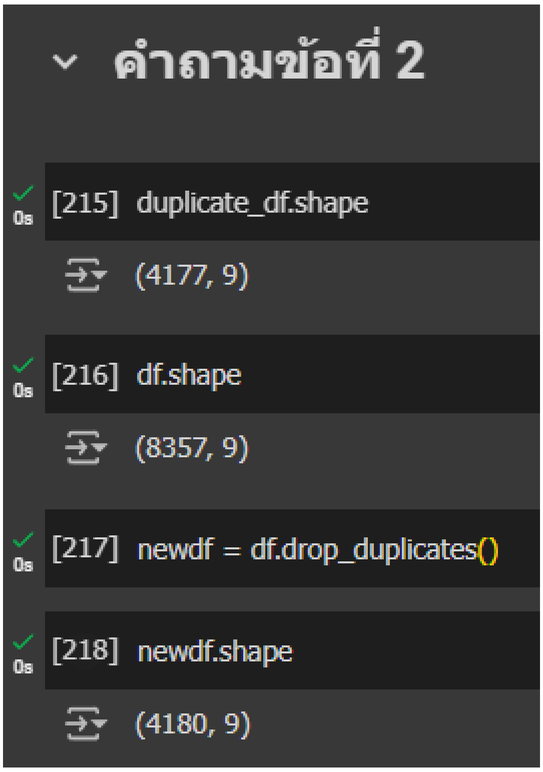


```
[213] duplicate_df  = df[df.duplicated()]
```

```
[214] duplicate_df.head()
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
4177	NaN	0.455	0.365	NaN	NaN	0.2245	0.1010	0.150	15.0
4178	M	0.350	0.265	NaN	NaN	NaN	0.0485	NaN	7.0
4179	F	0.530	0.420	0.135	NaN	0.2565	0.1415	0.210	9.0
4180	M	0.440	0.365	0.125	0.516	0.2155	0.1140	NaN	10.0
4181	I	0.330	0.255	0.080	0.205	0.0895	NaN	0.055	7.0

คำถามข้อที่ 2



คำถามข้อที่ 2

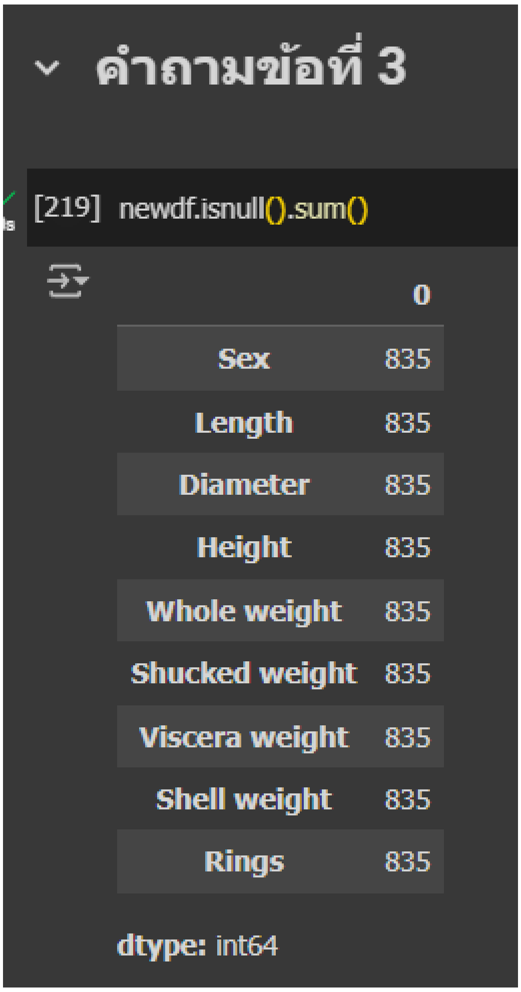
```
[215] duplicate_df.shape
(4177, 9)
```

```
[216] df.shape
(8357, 9)
```

```
[217] newdf = df.drop_duplicates()
```

```
[218] newdf.shape
(4180, 9)
```

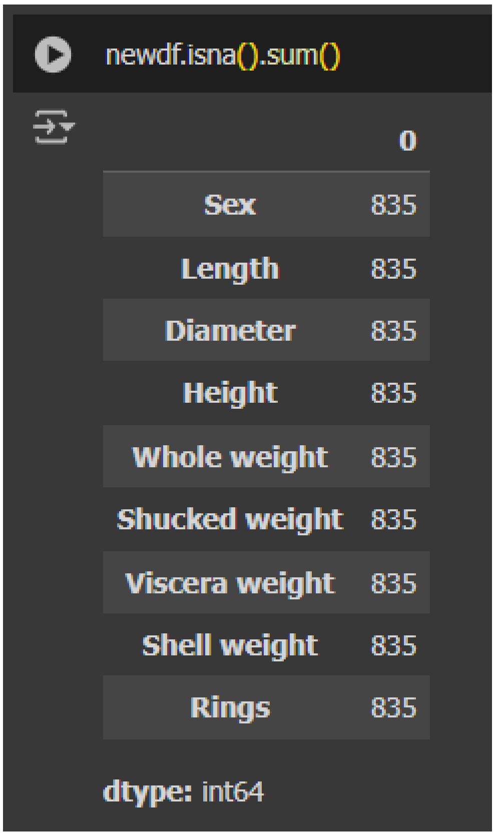
คำถามข้อที่ 3



คำถามข้อที่ 3

```
[219] newdf.isnull().sum()
```

```
                  0
           Sex  835
        Length  835
      Diameter  835
        Height  835
  Whole weight  835
Shucked weight  835
Viscera weight  835
  Shell weight  835
         Rings  835

dtype: int64
```



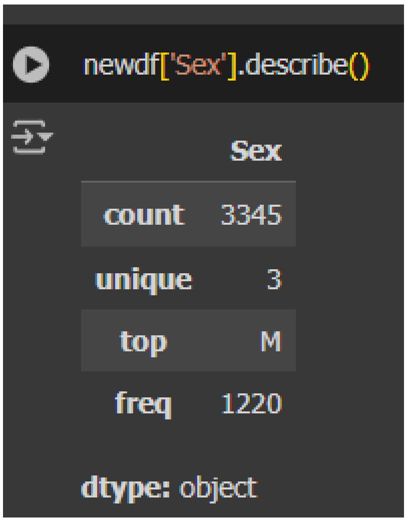
```
newdf.isna().sum()
```

```
                  0
           Sex  835
        Length  835
      Diameter  835
        Height  835
  Whole weight  835
Shucked weight  835
Viscera weight  835
  Shell weight  835
         Rings  835

dtype: int64
```

isna() กับ isnull() ไม่มีความแตกต่าง

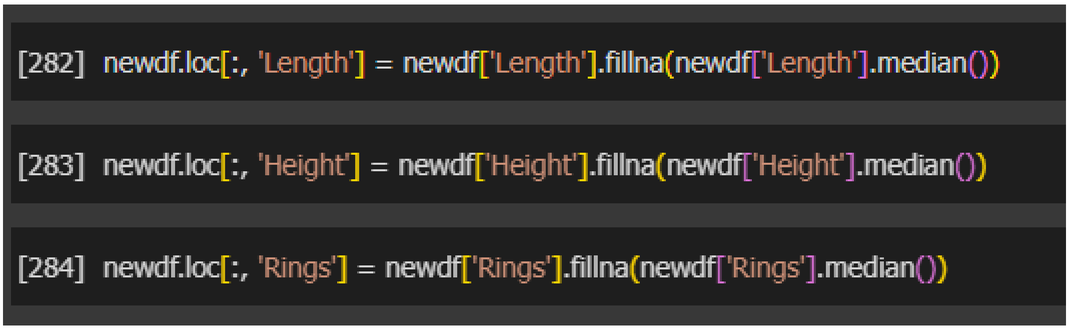
คำถามข้อที่ 4

คำถามข้อที่ 4



```
newdf['Sex'].describe()
```

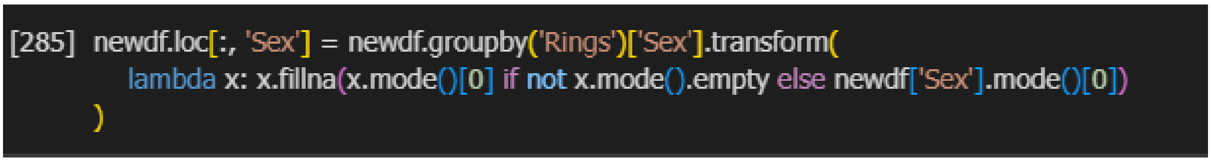
```
          Sex
 count   3345
unique      3
   top      M
  freq   1220

dtype: object
```



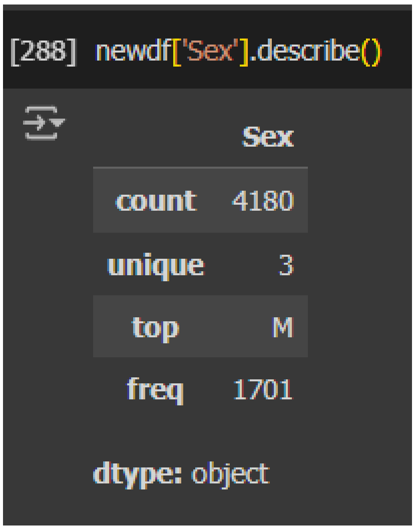
```
[282] newdf.loc[:, 'Length'] = newdf['Length'].fillna(newdf['Length'].median())
```

```
[283] newdf.loc[:, 'Height'] = newdf['Height'].fillna(newdf['Height'].median())
```

```
[284] newdf.loc[:, 'Rings'] = newdf['Rings'].fillna(newdf['Rings'].median())
```



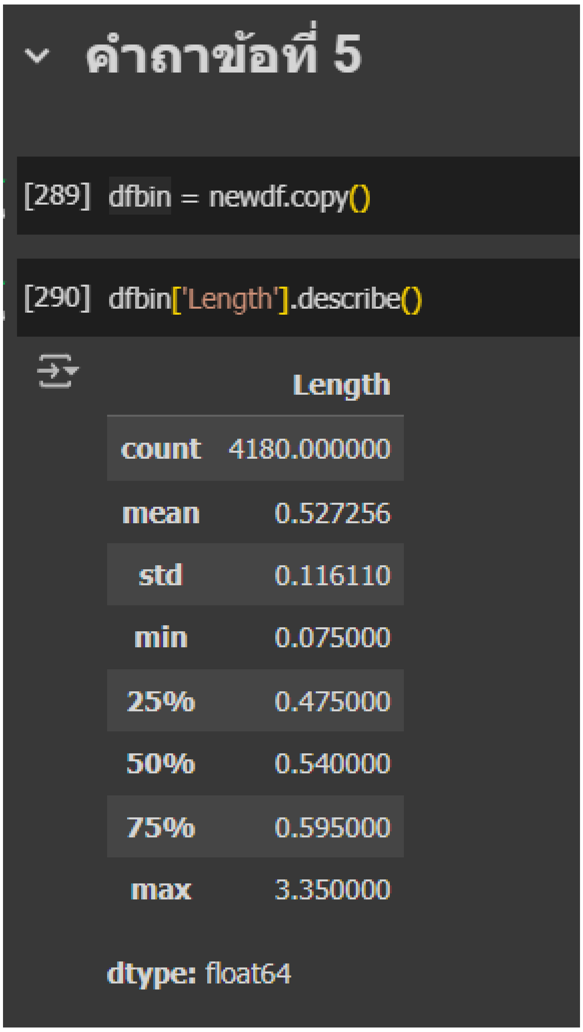
```
[285] newdf.loc[:, 'Sex'] = newdf.groupby('Rings')['Sex'].transform(
          lambda x: x.fillna(x.mode()[0] if not x.mode().empty else newdf['Sex'].mode()[0])
      )
```



```
[288] newdf['Sex'].describe()
```

```
          Sex
 count   4180
unique      3
   top      M
  freq   1701

dtype: object
```

คำถามข้อที่ 5

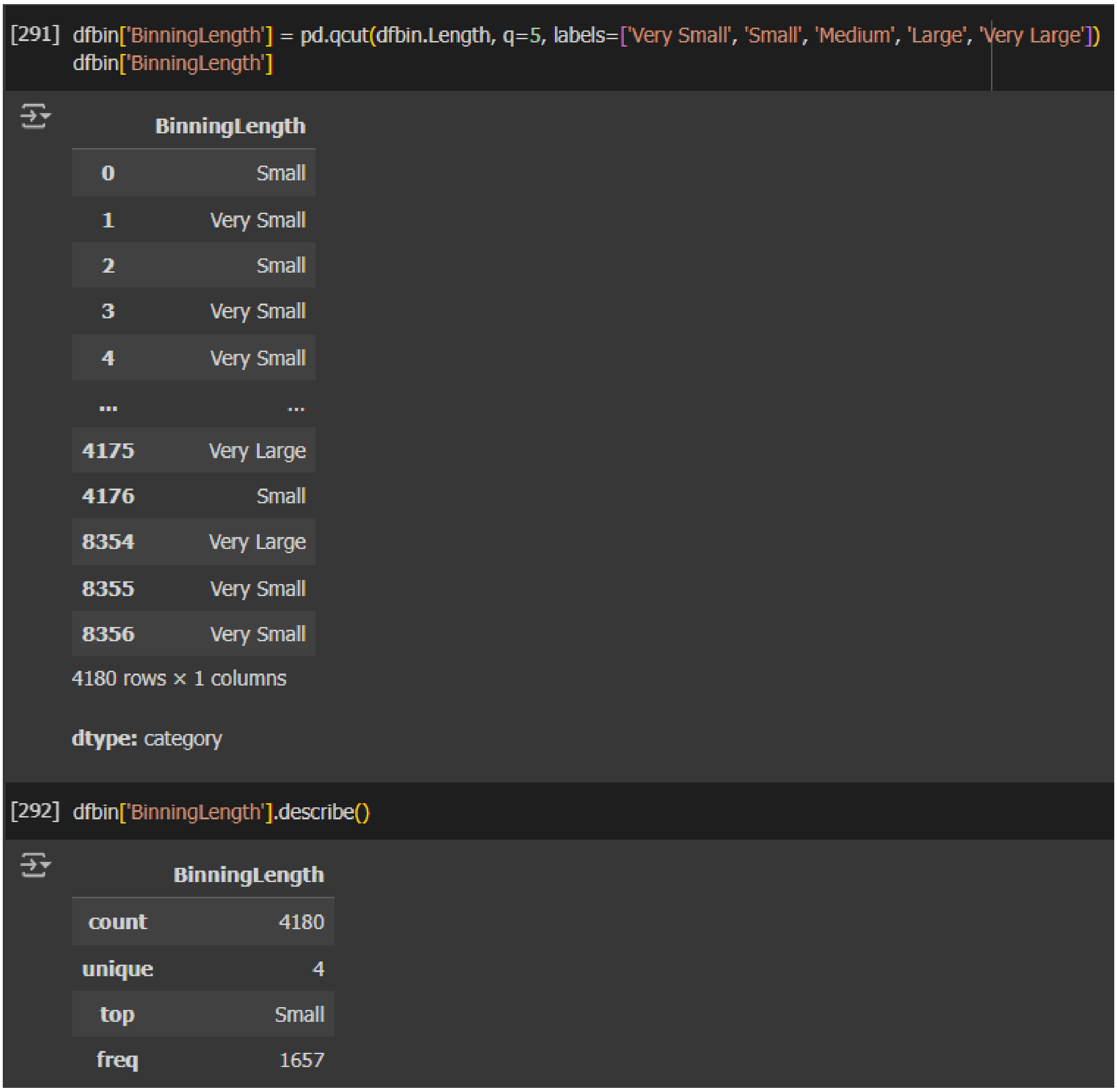


คำถาข้อที่ 5

```
[289] dfbin = newdf.copy()
```

```
[290] dfbin['Length'].describe()
```

```
             Length
count   4180.000000
mean       0.527256
std        0.116110
min        0.075000
25%        0.475000
50%        0.540000
75%        0.595000
max        3.350000

dtype: float64
```



```
[291] dfbin['BinningLength'] = pd.qcut(dfbin.Length, q=5, labels=['Very Small', 'Small', 'Medium', 'Large', 'Very Large'])
      dfbin['BinningLength']
```

```
        BinningLength
   0            Small
   1       Very Small
   2            Small
   3       Very Small
   4       Very Small
  ...             ...
4175       Very Large
4176            Small
8354       Very Large
8355       Very Small
8356       Very Small
4180 rows × 1 columns

dtype: category
```

```
[292] dfbin['BinningLength'].describe()
```

```
        BinningLength
 count           4180
unique              4
  top           Small
 freq            1657
```